DOCUMENT RESUME

ED 331 850 TM 016 353

AUTHOR Goodison, Jules M.
TITLE Training, Data Collection and Scoring of the 1990 Trial State Assessment.
PUB DATE Apr 91
NOTE 15p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7, 1991). Paper is based on the National Assessment of Educational Progress 1990 Technical Report and the Technical Report for the 1990 Trial State Assessment Program.
PUB TYPE Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; *Data Collection; Grade 8; Junior High Schools; *National Surveys; Program Implementation; *Scoring; *State Programs; Testing Problems; *Testing Programs; *Training
IDENTIFIERS National Assessment of Educational Progress; *Trial State Assessment (NAEP)

ABSTRACT

Procedures used in securing the cooperation of states and schools for the 1990 Trial State Assessment Program (TSAP) of the National Assessment of Educational Progress (NAEP) are described. These include procedures for: (1) the 1989 Field Test; (2) the training of local administrators and quality control monitors; and (3) scoring and processing answer booklets. Each participating jurisdiction selected a State Coordinator. At the local level, a Local Assessment Administrator was responsible for preparing for and conducting the assessment session in one or more schools. The NAEP hired and trained a State Supervisor for each state to work with the State Coordinator. Quality Control Monitors (four per state) were hired by the state collection subcontractor. The 1989 Field Test in 27 jurisdictions gave states an opportunity to rehearse for the 1990 TSAP in 40 jurisdictions. Processing and scoring for the 1990 TSAP involved 107,337 assessment booklets and 20,225 teacher and excluded student questionnaires. After collection, scoring was done by extensively trained readers of the 1990 National Program of the NAEP. Transcription of student responses into machine-readable form was achieved through data entry, validation, and resolution and storage. Three tables present study data. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED331850

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

JULES GOODISON

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Training, Data Collection and Scoring of The 1990 Trial State Assessment

by Jules M. Goodison
National Assessment of Educational Progress
Educational Testing Service
April 1991

Paper presented as part of the NAEP Symposium,"NAEP Reports On The State of the States," American Educational Research Association, 1991 Annual Meeting, Chicago, Illinois.

TM016353

BEST COPY AVAILABLE

This paper describes the procedures used in securing the cooperation of states and schools for the 1990 Trial State Assessment Program, including the 1989 Field Test, the training of the Local Administrators and Quality Control Monitors, and the scoring and processing of the answer booklets.

1. COOPERATION

The administration for the 1990 Trial State Assessment and 1989 Field Test for the Trial State Assessment both involved a collaborative effort between the participating states and the NAEP contractors. NAEP's responsibilities included:

- Selecting the sample of schools and students for each participating state;
- Developing the administration procedures and manuals;
- Training the personnel who would conduct the assessments; and
- Conducting an extensive quality assurance program.

Each jurisdiction volunteering to participate in the 1989 field test and the 1990 Trial State Assessment was asked to appoint a State Coordinator. In general, the State Coordinator was the liaison between NAEP staff and the participating schools. In particular, the State Coordinator was asked to:

- Gain the cooperation of the selected schools;
- Assist in the development of the assessment schedule;
- Receive the lists of all eighth-grade students from the schools;
- Coordinate the flow of information between the schools and NAEP;
- Provide space for the NAEP Supervisor to use when sampling;
- Notify local administrators about training and send them their manuals;
- Send the lists of sampled students to the schools.

This paper is based on the NAEP 1990 Technical Report and the Technical Report for the 1990 Trial State Assessment Program. (Both reports are now in the review process.)

At the local school level, a Local Assessment Administrator was responsible for preparing for and conducting the assessment session in one or more schools. These individuals were usually school or district staff and were trained by NAEP staff. The Local Administrator's responsibilities included:

- Receiving the list of sampled students from the State Coordinator;
- Identifying sampled students who should be excluded;
- Distributing assessment questionnaires to appropriate school staff;
- Notifying sampled students and their teachers;
- Administering the assessment session;
- Completing assessment forms; and
- Preparing the assessment materials for shipment.

NAEP hired and trained a State Supervisor for each state. The State Supervisors were responsible for working with the State Coordinators and overseeing assessment activities. Their primary tasks were:

- To make sure the arrangements for the assessments were set and local administrators identified;
- To schedule and conduct the local administrators' training sessions;
- To select the sample of students to be assessed; and
- To coordinate the monitoring of the assessment sessions and makeup sessions.

For the Trial State Assessment, Westat (the NAEP subcontractor for Field Data Collection) hired and trained four Quality Control Monitors in each state to monitor 50 percent of the assessment sessions. During the Field Test, the State Supervisors monitored all sessions.

1.1 The 1989 Field Test

In preparation for the 1990 Trial State Assessment Program, a field test of the forms, procedures, and booklet items was held in early 1989. The 1989 Trial State Assessment Field Test was designed to give states an opportunity to learn about their responsibilities and rehearse the data collection procedures envisioned for 1990 -- with a much smaller set of schools and students than would be included in 1990.

The 1989 Field Test was conducted from February 6 to March 3, 1989, in 23 states, the District of Columbia, and three U.S. Territories volunteering to participate. Table 1 lists the participating states and territories.

The intent of the Field Test was to give states and the NAEP contractors practice and experience with the proposed materials and procedures. For these purposes, it was decided that a sample of nine schools in each state would be sufficient.[1] For each state and the District of Columbia, WESTAT, the sampling and field operations contractor, selected nine pairs of schools clustered in two geographical locations. Because the sample for each state was so small, no attempt was made to make it representative at the state level; it was representative only at the national level. However, because of the small number of schools in the territories, all of their schools were asked to participate in the field test.

Each jurisdiction volunteering to participate in the 1989 State Field Test was asked to appoint a State Coordinator. In general, the State Coordinator was the liaison between NAEP staff and the participating schools. The State Coordinator was sent a list of the schools and requested to obtain the cooperation of one school from each pair. In a few rare instances, the State Coordinator could not obtain cooperation from either school in the pair. For these situations, NAEP provided a substitute school(s) which were asked to participate.

TABLE 1

Jurisdictions Participating
in the 1989 Field Test

States		
American Samoa	Michigan	Oregon
Arkansas	Minnesota	Pennsylvania
District of Columbia	Mississippi	South Carolina
Florida	Montana	Texas
Guam	Nebraska	Virginia
Illinois	Nevada	Virgin Islands
Kentucky	New Mexico	West Virginia
Louisiana	Ohio	Wisconsin
Maryland	Oklahoma	Wyoming

1 Because of the small number of schools in Guam and the Virgin Islands, all of their schools were asked to participate in the Field Tests.

All states were able to elicit the cooperation of the requisite number of schools. A total of 237 schools agreed to participate. Of the 237, 233 actually conducted assessments.

1.2 The 1990 Trial State Assessment

Thirty-seven states, the District of Columbia, and two territories volunteered for the 1990 Trial State Assessment. Table 2 lists the jurisdictions participating in the 1990 program.

A stratified random sample of approximately 100 schools was selected for each jurisdiction having that number of schools with the eighth grade.

Lists of the sampled schools were sent to the State Coordinators in July 1989, along with instructions describing their responsibilities in the assessment. Additional materials, such as forms to be used to list eligible students, NAEP reports, and descriptions of the schools' role in the assessment were sent by ETS and WESTAT in the Fall. State Coordinators were requested to determine the cooperation of the sampled schools and to notify Westat by November 1. Substitute schools were selected for those schools that had refused by that date. Table 3 provides the results of the State Coordinators efforts to gain the cooperation of the selected schools. In summary, ninety-four percent of schools in the original sample that could have participated in the assessment did so. An additional 100 substitute schools also agreed to participate. Thus, the participation rate, after substitution for refusals, was ninety-seven percent.

TABLE 2

State and Territories Participating
in the Trial State Assessment Program

STATES			
Alabama	Guam	Minnesota	Oklahoma
Arizona	Hawaii	Montana	Oregon
Arkansas	Idaho	Nebraska	Pennsylvania
California	Illinois	New Hampshire	Rhode Island
Colorado	Indiana	New Jersey	Texas
Connecticut	Iowa	New York	Virginia
Delaware	Kentucky	New Mexico	Virgin Islands
District of Columbia	Louisiana	North Carolina	West Virginia
Florida	Maryland	North Dakota	Wisconsin
Georgia	Michigan	Ohio	Wyoming

TABLE 3

Cooperation Rates

	Number	Percent
Schools in the original sample	3853	
Schools that were out of scope (closed, no eighth grade, not a regular school)	137	
Schools "eligible" to participate	3716	100.0%
Participating school	3496	94.1%
Non-participating schools (school or district refusal)	220	5.9%
Participating substitute schools	100	
Total participating schools	3596	96.8%

2. SUPERVISOR TRAINING

The supervisors selected to work on this project were recruited and hired during the summer of 1989. Supervisors were trained between September 15-19, 1989.
The training session focused on components of both the National Assessment and the Trial State Assessment. While the State Supervisors were trained separately in preparation for the Trial State Assessment, the entire group of Supervisors were trained on the general features of the Trial State Assessment, anticipating that the Assessment Supervisors would be questioned about the Trial State Assessment by district and/or school personnel. Also, there were presentations to all supervisors on the National Assessment since the State Supervisors would be conducting some introductory meetings for the National Assessment.

Topics on the training agenda included an overview of the National Assessment and the Trial State Assessment and the supervisors' responsibilities, a discussion of various reports from recent assessments, procedures for contacting districts and conducting introductory meetings, scheduling assessments within the PSUs, recruiting and training exercise administrators, procedures for drawing the sample of students, conducting assessments, preparation and distribution of questionnaires, and administrative forms and procedures. Also featured were practice exercises in sampling and filing out the various administrative forms.

A mock assessment session was held with the supervisors acting as "students." This included reading verbatim from one of the actual scripts to be used during an assessment and following prescribed procedures for distributing materials, reading directions, and recording the results of the assessment.

Additional training for the State Supervisors was held at the beginning of January 1990 when the second part of the State Supervisor training was conducted. The training took place between January 3 and January 7, 1990. The first three days focused on preparing the Supervisors for the training they would be conducting for the Local Administrators. During the final two days, the Quality Control Monitors (QCMs) also were at the training. The emphasis for the QCMs was on their responsibilities, primarily targeting the completion of the Quality Control Forms that were filled out for each assessment.

Shortly after their training, the Supervisors began conducting their training sessions for Local Administrators. In order to ensure uniformity in the training sessions, Westat developed a highly structured program involving a script for trainers, a video, and a training example to be completed by the trainees. The Supervisors were directed to read the script verbatim as they proceeded through the training. The importance of ensuring that all trainees received all of the same information dictated that such a script be used. The script was supplemented by the use of overhead transparencies, displaying the various forms that were to be used and enabling the trainer to demonstrate how they were to be filled out.

The videotape was developed by Westat to provide background for the study and to

simulate the various steps of the assessment that would be repeated by the LA's. The portions of the video depicting the actual assessment were taped in a classroom with students in attendance so that a close representation of an actual assessment could be provided. The video was divided into five sections, with breaks for review by the trainer and practice for the trainees.

The final component of the presentation was the "Training Example for Local Administrators." This consisted of a set of exercises keyed to each part of the training package. A portion of the video and part of the script were presented, and then exercises related to that material were completed before going on the next subject.

Prior to attending the training session, each Local Administrator received the Manual for Local Administrators provided by NAEP to the State Coordinator. This manual was a comprehensive document which explained every step of assessment procedures.

The entire training session generally ran for about five to six hours. In all, a total of 356 LA training sessions were held, (an average of nine per state), with 3463 LA's being trained.

3. PROCESSING AND SCORING

This section details the receipt, processing and final disposition of the 1990 Trial State Assessment materials. The scope of the effort required to process the 1990 Trial State Assessment materials is evidenced by the following:

- 107,337 assessment booklets and 20,225 teacher and excluded student questionnaires were received and processed.
- Approximately 2.2 million double-sided pages from test booklets and questionnaires were optically scanned.
- 3,538 school questionnaires were manually key-entered and verified.

3.1 Processing Assessment Material

The materials from each session were to be returned to NCS (processing subcontractors) in the same box in which it was originally mailed. It was the responsibility of the Local Administrator in the unmonitored schools and the Quality Control Monitor in the monitored schools to re-package the items in the proper order, complete all paper work and return the shipment via the U.S. Postal Service, utilizing the postage paid label provided.

With approximately 4,000 individual shipments arriving over a four week period, it was necessary to devise a system that would quickly acknowledge receipt of a school's

material. The label used to ship the materials to the school or administrator contained a bar code which indicated the school number and the project number. When the shipment arrived at NCS, the bar code was read and the shipment forwarded to the receiving area. The file was then transferred to the mainframe via a PC link and a computer program was used to apply the shipment receipt date to the appropriate school within the system. This provided current status of shipments received regardless of any processing backlog. Any shipment that was not received within seven days of the scheduled assessment was flagged on a report that was issued to Westat. The status of the administration was checked and in some cases a trace was initiated on the shipment.

Receiving personnel also checked the shipment to verify that the contents of the box matched the school and session indicated on the label. Each shipment was checked for completeness and accuracy, regardless of whether it was a monitored or unmonitored session.

The materials were checked against the packing list to verify that all materials were returned. If any discrepancies were found, an alert was issued. If all assessment instruments were returned, then processing continued. Quantities of scientific calculators were in short supply; therefore, during the first two weeks of the assessment, calculators were taken from the incoming shipments and returned to the packaging area to be included in other shipments for the last weeks of testing.

Each booklet and Excluded Student Questionnaire was verified against the administration schedule. This included verification of a all counts of booklets returned and the matching of information on the front cover of the booklets to that on the administration schedule. If any discrepancy was discovered, an alert was issued.

After the contents of the shipment had been identified and verified, the packing list information was entered into the system. That information included school number, session number, counts of the number of used and unused booklets, and makeup status. If a makeup session was expected, an information alert was issued to facilitate tracking. The control counts were used for verification of processing counts.

If quantities and individual information matched, the booklets were organized into work units and batched by session. Each session was assigned a unique batch number composed of a session identified and the school number. A scannable session header was coded and placed on top of the stack of documents. All student documents were forwarded to professional scoring, other documents to a key entry activity, others directly to a machine scanning function, and others to appropriate record filing systems.

The Excluded Student Questionnaires and Teacher Questionnaires were compared to the Roster of Questionnaires and the administration schedule to verify demographic information. Some questionnaires may not have been available for return with the shipment. These were returned to NCS at a later date in an envelope provided for that

purpose.

The School Characteristics and Policies Questionnaire was compared to the Roster of Questionnaires and the school number was verified to match all other materials in the shipment. As with the other questionnaires, this document may not have been returned with the shipment and could also be returned in the supplemental envelope. There was no additional effort made to collect any unreturned questionnaires.

Unlike the National Assessment, absent students were not accounted for by use of an Absent Student Form. Absent students were assigned a test booklet. To indicate an absence, the "A" bubble in the Administration Code column on the front cover of the booklet was gridded. The booklet was then placed into processing along with assessed booklets to maintain session integrity.

The packing list was used by the schools to account for all materials received from and returned to NCS were indicated. Also indicated was whether a makeup session was to be held, the date of scheduled makeup, number of students involved, and quantities of materials being held for later return.

3.2 Professional Scoring

The student assessment booklets were forwarded to the professional scoring area once all appropriate materials were received from each school. Like the National Assessments, the 1990 Trial State Assessment included open-ended items. Open-ended and multiple choice items were administered in scannable assessment booklets that were identical to the mathematics booklets used in the National Assessment at the eighth grade level. Scores for the open-ended items in these samples were gridded in ovals at the bottom of the pages on which the items appeared.

The scoring of the 1990 Trial State Assessment was conducted simultaneously with the scoring of the mathematics portion of the 1990 National Program and the same readers scored both the National Assessment and the Trial State Assessment.

3.2.1 Description of Scoring

Each of the 35 open-ended items included in the 1990 Trial State Assessment had a unique scoring guide that identified the range of possible scores for the item and defined the criteria to be used in evaluating students' responses.

Eighteen items were categorized as a right/wrong while 17 items included categories of specific correct and incorrect responses. For the items scored as "right/wrong", a correct response was coded a score of "8" and an incorrect response was coded a score of "1". Items with more than one correct response were given a score of "7" for the alternate response.

Various incorrect responses were also tracked with separate score points. The incorrect responses were assigned a score point from one to five to capture information on the specific type of error the student was making.

The readers were organized into five teams of twelve readers plus one team leader. The five team leaders reviewed discrepancies between readers and reviewed decisions regularly so that all readers approached a specific problem in the same manner.

3.2.2 Training

Before the training program began, the team leaders worked with NAEP Test Development staff to prepare training sets, (or sets of sample responses to accompany the scoring guides), and to refine the scoring guides for newly developed items.

The training was conducted by ETS mathematics specialists with assistance from the five team leaders. Training involved explaining the item and its scoring guide and discussing responses that were representative of the various score points in the guide. When this process was competed, the readers scored and discussed 5 to 20 randomly selected "practice papers" for each item, depending on the complexity of that item. Each reader on a team scored all the open-ended items in each of approximately 12 bundles of booklets, with an average of 27 booklets per bundle. During this practice, discussion sessions were held to review responses that received a wide range of scores. Once the practice session was completed, the formal scoring process began. The initial training was completed in one week.

As a follow-up to the training program, notes on various items were compiled for the readers of each group for their reference and guidance throughout the scoring process. In addition, short training sessions were conducted when the team leaders determined by reviewing discrepancies that certain items were causing difficulties for the scorers. The team leaders also consulted with individual readers as the scoring progressed. When a reader's score was judged to be discrepant with that of another reader, the supervisor discussed the response and its score with that reader.

3.2.3 Paper Flow

Twenty percent of all open-ended responses were to be scored by a second reader to obtain statistics on inter-reader reliability. The reliability information is used as a way to monitor the capabilities of particular readers and the uniformity across readers about each task.

The addition of separate reliability scoring sheets, along with the introduction of performing the reliability scoring step first served two purposes. It ensured that all reliability scoring was truly "blind", uninfluenced by any score already given, and it expedited the scoring process by eliminating the need to mask the scores assigned by the primary

reader.

When the batches had been scored for reliability and had been scored by the primary scorer, the Team Leader verified that all batches had been completed and retrieved the corresponding reliability scoring sheets from the folder in which they were held. Upon completion of scoring for a batch, the scannable booklets and separate scoring sheets were forwarded to the next workstation to be "super" batched for further processing.

Inter-reader reliability was calculated for all items throughout the scoring period. Any particular item was read twice at least 9200 times. The average inter-reader reliability for all items was 96.9%.

4. DATA TRANSCRIPTION SYSTEMS

The transcription of the student response data into machine-readable form was achieved through the use of three separate systems: data entry, (scanning or key entry), validation, (pre-edit), and resolution.

4.1 Data Entry

The data entry process is the first time that booklet level data were directly available to the computer system. Depending on the document, one of two methods was used to transcribe data to a computerized form. The data on scannable documents was collected using NCS optical scanning equipment. Non-scannable material was keyed through an interactive on-line system. In both cases, the data was edited and questionable data were resolved before further processing.

In order to ensure data integrity, edit rules are applied to each scanned data field. This procedure validates each field and reports all problems for subsequent resolution. After each field is examined and corrected, the edit rules are re-applied for final verification.

4.2 Scanning

After the professional scoring, the scannable documents (the student booklets, Excluded Student Questionnaires, and Teacher Questionnaires) were then transported to a slitting area where the folded and stapled spine was removed from the document. Scanning operations were performed by NCS's HPS Optical Scanning equipment. The optical scanning devices and software used at NCS allows a complete mix of NAEP scannable materials with no special grouping requirements. However, for manageability and tracking purposes, student documents, Excluded Student Questionnaires and Teacher Questionnaires were batched separately. In addition to the capture of scannable responses, the bar code identification numbers used to maintain process control were also decoded and transcribed to the NAEP computerized data file.

4.3 Data Validation

The data entry and resolution system used for the Trial State Assessment Program is the same system used for the National Assessment Program. The system is able to process materials from three age groups simultaneously, three assessment types, one absent form and five questionnaires submitted to the system from scannable and non-scannable media. The use of batch identification codes -- comprising the school and session codes as well as the batch number, facilitated the management of the system.

When the entire document has been processed, the completed string of data is written to the data file. When the next session header record is encountered, the program repeats the same processes for that session. When the program encounters the end of a file, it closes the data set, and generates an edit listing.

Accuracy checks were performed on each batch processed. Every five hundredth document of each book form was printed in its entirety, with a minimum of one document type per batch. This record was checked manually, item by item, with the source document for errors.

4.4 Storage of Documents

Once the editing process had been successfully completed on the batches, the documents were sent to a warehouse for storage. The storage location of all documents was recorded on the inventory control system and stored for later retrieval. Unused materials were sent to temporary storage until the completion of an assessment and acceptance of the data tape, at which time they were destroyed.